# Vers une modélisation continue de la structure prosodique : le cas des proéminences syllabiques

Mathieu Avanzi, Anne Lacheret, Nicolas Obin, Bernard Victorri

*French Language Studies* **21** (2011), 53–71, © Cambridge University Press
doi:10.1017/S0959269510000517

# *Vers une modélisation continue de la structure prosodique: le cas des proéminences syllabiques*

MATHIEU AVANZI, *ANNE LACHERET-DUJOUR, +NICOLAS OBIN et °BERNARD VICTORRI

*Université de Neuchâtel et Université Paris-Ouest,*
**Université Paris-Ouest – MODYCO*
*+IRCAM, Paris*
*°Lattice/ENS, Paris*

(Received March 2010; revised September 2010)

RÉSUMÉ

L'objectif de cet article est de présenter un outil développé en vue de modéliser semi-automatiquement la structure prosodique du français. Sur la base d'un alignement en phonèmes, notre système procède à la détection des syllabes proéminentes en prenant en considération des critères acoustiques basiques tels que la fo, la durée et la présence de pauses. À partir des mesures ainsi prises, le système attribue un degré de proéminence à chacune des syllabes identifiées comme saillante. Nous illustrons ensuite les résultats de l'analyse d'extraits du corpus PROSO_FR. Plus précisément, nous comparons l'analyse prosodique de phrases que l'on pourrait faire avec les règles traditionnelles de la phonologie prosodique avec l'analyse conduite par notre logiciel. Nous discutons ainsi de trois règles: la règle de dominance droite, la règle de clash accentuel et la règle des sept syllabes.

## 1. INTRODUCTION

### *1.1 La grille métrique*

Dans la communauté des chercheurs en prosodie, il y a accord sur le fait que mettre au jour le profil prosodique d'un énoncé dans une langue revient notamment à dégager sa structure rythmique[1] (Dell, 1984), structure que l'on a coutume de représenter, depuis Prince (1983), sous la forme d'une *grille métrique*:

[1] La structure rythmique n'est qu'une représentation partielle de la structure prosodique, car elle ne permet pas de rendre compte du *profil intonatif* de l'énoncé, c'est-à-dire de la forme des contours mélodiques associés aux syllabes finales de groupe accentuels. Des systèmes de transcription automatique tels que Momel (Hirst et Espesser, 1993) ou Prosogramme (Mertens, 2004) ont été conçus à dessein pour traiter de tels aspects (*cf.* pour une présentation de ces systèmes Delais-Roussarie et Yoo, ce volume). Nous nous servons de la grille métrique comme d'un outil de représentation, sans forcément adhérer aux présupposés théoriques qui lui sont associés (cf. Astésano, 2001).

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | * |
| | | | | | * | | | | * | | | | * | | * |
| | * | | * | | * | | | | * | | | | * | | * |
| * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * |
| il | dɔn | lə | sɑ̃ | ti | mɑ̃ | da | vwaʁ | se | de | a | la | pʁe | sjɔ̃ | ɑ̃ | bjɑ̃t |
| il | donne | le | sentiment | | | d'avoir | | cédé | | à | la | pression | | ambiante | |

Figure 1. *Grille métrique étiquetée associée à l'énoncé* il donne le sentiment d'avoir cédé à la pression ambiante *(For-Lec)*.

Les syllabes de l'énoncé analysé dans la figure 1, transcrites en API dans une couche dédiée (une « tire » d'annotation), sont toutes potentiellement accentuables, et donc accompagnées d'au moins un astérisque. Aux syllabes effectivement accentuées sont associées au moins deux astérisques. Le nombre d'astérisques associé à la syllabe est relatif à sa force perceptive, *i.e.* à son degré de proéminence (Halle et Vergnaud, 1987).

### *1.2 La méthode traditionnelle pour construire la grille*

Traditionnellement, les phonologues travaillant sur le français proposent, pour mettre au jour la grille métrique associée à un énoncé donné, de la « dériver » à partir d'indices syntactico-sémantiques et informationnels, tout en considérant les contraintes métriques propres à cette langue. Nous n'avons pas ici la place d'entrer dans les détails des règles existantes pour prédire la localisation et la force des accents dans la phrase (voir cependant notre partie 3 où nous commentons trois exemples de règles métrico-syntaxiques traditionnelles). Aussi nous contenterons-nous de dire qu'en ce qui concerne le français, les spécialistes partent de l'idée selon laquelle le système accentuel du français est '*bipolaire*' (*cf.* Di Cristo, ce volume). Un accent, dit '*final*' (ou '*primaire*' dans certaines terminologies), obligatoire, frappe la dernière syllabe masculine (syllabe dont le noyau n'est pas un schwa) des groupes de mots contenant au moins un morphème lexical et les clitiques qui en dépendent. L'autre accent, dit par opposition '*non-final*' (ou '*secondaire*') et dont la présence est motivée par des contraintes de régulation rythmique, essentiellement, frappe n'importe quelle autre syllabe du groupe ponctué à sa droite par l'accent primaire. Il est très difficile de prédire la place de ces deux accents en français: si dans la théorie les indices morphosyntaxiques de segmentation en *chunks* peuvent aider à prédire les syllabes porteuses d'un accent, dans la pratique les choses ne sont pas si simples, des contraintes rythmiques liées au débit des locuteurs, à la formation de groupes métriquement équilibrés, etc., venant contraindre les prédictions établies par la morphosyntaxe (*cf.* Lacheret-Dujour et Beaugendre, 1999 pour un aperçu exhaustif de l'ensemble de ces règles et des références sur le sujet). Pour déterminer le degré de proéminence d'une syllabe accentuée, les chercheurs prennent en considération

le degré d'enchâssement syntaxique du constituant syllabique en question (plus un constituant est enchâssé dans la phrase, moins son degré de proéminence est important; de deux constituants dominés par le même nœud, c'est celui de droite qui porte la proéminence la plus forte (*cf.* Dell, 1984 et les contributions réunies dans ce volume pour des revues) et, plus récemment, les indices informationnels, relatifs à l'articulation fond/focus et au degré de saillance des topiques (la frontière droite du focus est la plus forte de l'énoncé, un topique déjà introduit n'est pas nécessairement fortement accentué, etc. (*cf.* Delais-Roussarie, 2005; Delais-Roussarie et Post, 2008). Considérons à titre d'illustration l'exemple ci-dessus. Chacune des syllabes terminales des mots lexicaux délimite un '*groupe accentuel*'. Ainsi, l'énoncé est formé de cinq groupes accentuels *(il donne) (le sentiment) (d'avoir cédé) (à la pression) (ambiante)*[2]. Seul le second groupe est porteur d'un accent initial (*sentiment*). La force des proéminences finales qui ponctuent les groupes permet de distinguer trois niveaux dans la structure prosodique (trois niveaux de 'phrasé' dirait Di Cristo, ce volume): celui du '*syntagme accentuel mineur*', celui du '*syntagme accentuel majeur*' et celui du '*syntagme intonatif*' (nous reprenons les termes en usage dans les travaux de Selkirk, 2005). Le syntagme intonatif est le plus haut dans la hiérarchie: il englobe un syntagme accentuel majeur, qui lui englobe à son tour au moins un syntagme accentuel mineur (voir Portes et Bertrand, ce volume). Dans l'exemple ci-dessus (fig. 1), l'ensemble de la phrase forme un seul et unique syntagme intonatif, composé de quatre groupes accentuels majeurs (syllabes alignées avec des colonnes formées de 3 astérisques au moins) et de cinq groupes accentuels mineurs (2 astérisques au moins sur les syllabes finales).

### *1.3 Objectifs de cet article*

Mettre au jour de façon semi-automatique la structure rythmique des énoncés que l'on trouve dans les corpus de français parlé constitue un enjeu majeur pour les chercheurs. Pour autant, dans une perspective descriptive, la démarche décrite sous §1.2. est extrêmement coûteuse et difficile à appliquer en l'état. Elle nécessite en effet que l'on contrôle un certain nombre de facteurs, et qu'on ait une connaissance fine de la façon dont ils interagissent, ce qui est loin d'être le cas aujourd'hui. Concrètement, on ne sait pas encore très bien dans quels contextes précis les contraintes rythmiques peuvent contrebalancer les prédictions morphosyntaxiques: une frontière de constituant syntaxique majeur (typiquement un ajout à la phrase matrice, tel qu'un constituant extraposé ou disloqué) peut-elle être effacée si les contraintes rythmiques l'autorisent? Ce genre de segment, censé revêtir un statut informationnel bien précis (jouant le rôle de topique ou de post-focus), est-il obligatoirement entouré de frontières prosodiques fortes ? Les réponses à de telles questions sont loin d'aller de soi, et si on a pu apporter il y a quelques années des éclairages intéressants dans le cadre de la théorie de l'optimalité

[2] Le verbe *avoir* ne constitue pas un mot attracteur d'accent final dans cette position (Delais-Roussarie, 2008: 66).

(Delais-Roussarie, 2005), des travaux plus récents portant sur l'analyse de constructions à l'interface de la syntaxe et du discours dans la parole spontanée ont montré que les contraintes rythmiques sont en fait plus importantes que ce que l'on avait pu croire jusqu'alors (Dehé, 2009; Avanzi, à par. a et b), la langue parlée non-préparée non-lue s'affranchissant parfois des règles de bonne formation propres à la parole lue (Guaïtella, 1997; Astésano 2001: 26; Lacheret-Dujour, 2003; Simon, 2004). Pour ces deux raisons, nous avons préféré, plutôt que de dériver la structure prosodique des énoncés de français spontané à partir d'indices extra-phonologiques (syntaxiques, sémantiques et informationnels) et métriques, suivre une approche inverse en vue de modéliser semi-automatiquement la structure prosodique des énoncés du français. Cela veut dire que nous avons cherché à faire émerger de la substance les corrélats acoustiques des proéminences accentuelles sans préjuger au départ des contraintes fonctionnelles qui pourraient les sous-tendre afin de mettre au point un algorithme capable de les identifier et de les catégoriser automatiquement par la suite. Dans la section suivante, nous présentons la démarche que nous avons suivie pour ce faire. Le corpus sur lequel nous appuyons nos analyses est le corpus PROSO_FR, présenté dans l'introduction de ce volume (cf. Avanzi et Delais-Roussarie, ce volume).

## 2. Présentation de l'algorithme d'identification et de catégorisation automatique des proéminences

Une proéminence est en général définie comme une entité syllabique qui se détache de son environnement comme une figure sur un fond en vertu d'un certain nombre de paramètres acoustiques et perceptifs (Terken et Hermes, 2000 et Avanzi, à par. b: chap. III, pour une synthèse et des références). D'après cette définition, la notion de proéminence est proche de celle d'*accent*, mais elle s'en distingue dans la mesure où elle correspond davantage à un fait phonétique (de substance) que phonologique (grammatical), et qu'elle ne préjuge pas de la place et d'une fonction *a priori* de l'événement dans la phrase. Elle est donc relativement neutre d'un point de vue théorique, et en ce sens, synonyme de la notion de *battement* (*beating*), à la base de la construction des grilles métriques de Prince (1983).

### *2.1 Identification automatique des proéminences*

Nous présentons ici la méthode que nous avons suivie pour mettre au point un détecteur de proéminence robuste. Outre la définition très générale proposée ci-dessus, nous sommes partis de l'idée que (i) la proéminence était syllabique; (ii) que la proéminence était un phénomène local (donc qu'il fallait définir une fenêtre restreinte pour identifier les variations prosodiques); (iii) que si les paramètres impliqués dans la perception de la proéminence étaient nombreux, les variations de fréquence fondamentale et de durée étaient les plus importants pour le français (*cf.* déjà Delattre, 1938, ainsi que Goldman et al., 2007).

*2.1.1 Calcul des paramètres acoustiques*

Afin d'identifier les syllabes proéminentes, il nous a fallu dans un premier temps délimiter une fenêtre contextuelle pertinente à l'intérieur de laquelle l'algorithme allait procéder au calcul des variations prosodiques significatives. Dans une étude précédente (Avanzi, Lacheret-Dujour et Victorri, 2010), nous avions montré qu'une fenêtre contextuelle trop étroite tendait à la sur-détection de proéminences, alors qu'une fenêtre contextuelle trop large tendait à la sous-détection. En conséquence, une fenêtre telle que le groupe accentuel, unité de taille intermédiaire, semblait un bon compromis. Pour nous rapprocher du syntagme accentuel sans faire intervenir de critères grammaticaux (le syntagme accentuel, défini comme un regroupement prosodique minimal de syllabe ponctué par un accent primaire, constitue un ensemble formé d'un mot lexical et des clitiques qui gravitent autour, *cf. supra* §1.2.), nous avions proposé de prendre comme contexte pour le calcul des variations prosodiques les trois syllabes qui précèdent et les trois syllabes qui suivent la syllabe dont on cherche à déterminer si elle est proéminente ou non, ce qui revient à travailler dans un intervalle de sept syllabes (sept étant le nombre de syllabes du syntagme accentuel maximal selon Wioland, 1985; Martin, ce volume).

La procédure de détection de proéminence a été implémentée sous Matlab par B. Victorri dans une interface logicielle nommée Analor (Avanzi, Lacheret-Dujour et Victorri, 2008 et 2010). Dans la figure 2, nous voyons (i) l'évolution de la f0 (en traits noirs). Cette dernière peut être mesurée en demi-tons (en filigrane, la distance entre deux lignes fines vaut 1 demi-ton, la distance entre deux traits épais une quarte (4 demi-tons)) ou en hertz (les valeurs numériques sont affichées sur la gauche); la durée des segments étiquetés est donnée en millisecondes au dessus de la bande dans laquelle évolue la courbe de f0. Les chiffres plus foncés au-dessus indiquent le temps en secondes du segment par rapport à l'enregistrement du fichier total. Les différentes couches d'alignement, importées directement depuis les fichiers d'alignement au format textgrid (Praat) sont affichées en dessous de cette bande, en l'occurrence, de haut en bas: les phonèmes et les syllabes en alphabet SAMPA, et la tire des mots graphiques (*cf.* Avanzi et Delais-Roussarie, ce volume).

Pour identifier les saillances syllabiques, nous avons décidé de considérer quatre variables (*cf.* figure 2). Pour chacune des syllabes de l'énoncé, (i) le logiciel procède au calcul de la moyenne de hauteur relative de la syllabe-cible par rapport à la moyenne de hauteur de tous les points de f0 des trois noyaux vocaliques des syllabes qui précèdent ($S_{-3}$, $S_{-2}$, $S_{-1}$), et des trois qui suivent ($S_{+1}$, $S_{+2}$, $S_{+3}$); (ii) il mesure également la moyenne de durée de l'ensemble des syllabes dans ce même empan[3]; (iii) il regarde ensuite si le noyau vocalique de la syllabe est porteur d'un glissando

[3] Sur ce point, idéalement, un modèle de durée des syllabes fondé sur les propriétés syllabiques intrinsèques et les variations locales de débit serait le plus approprié. Cependant, un tel modèle nécessite encore d'être développé et testé pour le français parlé non lu (pour une approche sur la lecture, *cf.* Obin *et al.*, 2008), nous avons seulement pris en compte le nombre de phonèmes dans la syllabe comme facteur de relativisation, excluant ainsi d'une certaine façon le biais du poids syllabique: p. ex. une syllabe composée de cinq phonèmes ([ʃkʁwa]) sera par nature plus longue qu'une syllabe mono-phonémique ([i]).

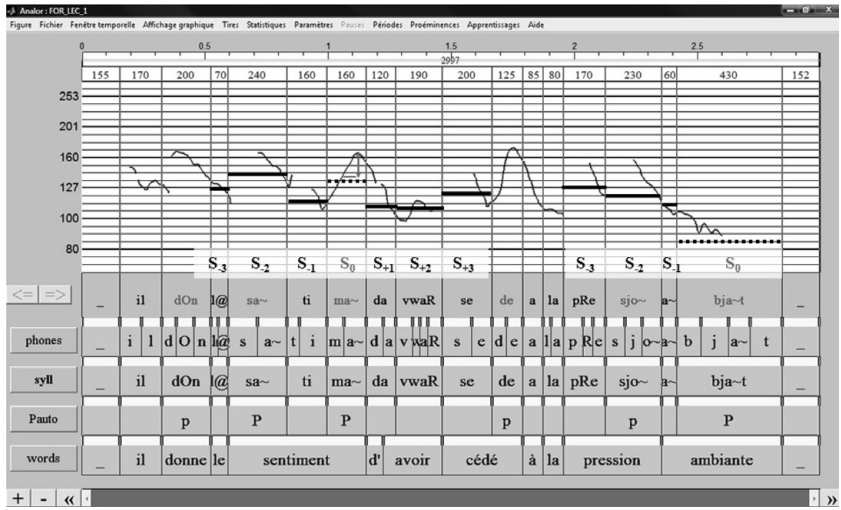


Figure 2. *Copie d'écran* ANALOR. *Illustration de la détection de proéminence.* Analyse de l'énoncé: il donne le sentiment d'avoir cédé à la pression ambiante *(*FOR-LEC*).*

montant, et calcule l'amplitude de ce glissando, le cas échéant (sur l'exemple ci-dessus, cela ne concerne que la syllabe [mɑ̃], les noyaux des autres syllabes ne portant pas de glissando montant); (iv) enfin, il prend en compte la présence d'une pause silencieuse qui suit directement la syllabe, peu importe la durée de cette pause (l'étiquetage des pauses silencieuses ayant été effectué lors de la phase de transcription et d'alignement semi-automatique du signal de parole, il n'y a aucune chance pour que les silences pré-occlusifs ou autre « fausses pauses » soient considérés comme des marques de fin de groupe). Alors que les valeurs de fo sont données en demi-tons, les valeurs de durée sont sans unité (une syllabe est *n* fois plus longue ou plus courte en moyenne que ses voisines). En outre, on remarquera que la relativisation peut être bloquée dans certains contextes: quand la syllabe est suivie d'une pause silencieuse, comme c'est le cas de la syllabe finale de *ambiante* dans l'exemple ci-dessus, les valeurs de fo et de durée relatives sont calculées par rapport aux trois syllabes qui précèdent.

*2.1.2 Algorithme*

Pour chacun de ces quatre paramètres, la méthode pour fixer les valeurs des seuils consistait à réaliser un apprentissage supervisé utilisant un corpus préalablement annoté, l'objectif étant d'optimiser une mesure de la performance du système en comparant les résultats obtenus sur ce corpus avec les données des annotateurs humains. Le corpus que nous avons employé pour ce faire est le corpus C-PROM. Cette base de données, dont la procédure de constitution et le contenu sont décrits en détail dans Avanzi, Simon, Goldman et Auchlin (2010), se compose d'un ensemble d'enregistrements d'une durée de 70 minutes, comprenant

sept paquets d'enregistrements de 10 minutes, chacun constituant un genre de discours spécifique[4], et comprenant des locuteurs francophones de France métropolitaine, de Suisse romande et de Belgique wallonne. Il a été transcrit et aligné semi-automatiquement en phonèmes, syllabes et mots graphiques. Les syllabes proéminentes et les syllabes dysfluentes (annotées pour ne pas gêner les relativisations des durées syllabiques, *cf.* Avanzi, Simon, Goldman et Auchlin, 2010 et Avanzi, à par. b pour les détails) ont été identifiées sur la base d'une analyse perceptive conduite par deux experts en prosodie.

L'algorithme, utilisé pour l'apprentissage, est basé sur une recherche aléatoire locale, à pas décroissant, dans l'espace des paramètres à partir d'une valeur initiale pertinente (l'explication complète est donnée dans Avanzi, Lacheret-Dujour et Victorri, 2010 et Avanzi, à par. b). Cet algorithme d'apprentissage n'étant efficace que si les valeurs initiales des paramètres sont suffisamment proches des valeurs optimales, les valeurs initiales pour l'entraînement ont été fixées sur la base d'une expertise linguistique préalable. À partir de travaux sur la perception des variations prosodiques, nous avions proposé les valeurs suivantes. Pour la hauteur relative, le seuil a été fixé à 1.5 dt (Rossi *et al.*, 1981; Rietveld et Gussenhoven, 1985; 't Hart *et al.*, 1990). Quant aux valeurs de durée relative, nous avons fixé ce seuil à 1.5 (ce qui revient à considérer qu'un allongement est significatif s'il dépasse 50% de la durée moyenne des syllabes environnantes), conformément aux observations de Rossi *et al.* (1981) et Lacheret-Dujour (2003). Enfin, en ce qui concerne le seuil initial de glissando, nous avons suivi Rossi *et al.* (1981) et Mertens et d'Alessandro (1995) et fixé le seuil à 3dt (demi-tons). Considérant que la présence d'une pause silencieuse est un indice fort de fin de groupe (Lacheret-Dujour et Beaugendre, 1999), quelle que soit la durée de cette dernière, nous n'avons pas eu à entraîner ce seuil (*cf.* Avanzi, Lacheret-Dujour et Victorri, 2008 pour une justification).

Après entraînement sur la totalité du corpus C-PROM, la performance du logiciel était de 80.96% de f-mesure, performance tout à fait honorable, surtout quand on sait que le taux d'accord entre deux annotateurs humains s'élève à 82.6% de f-mesure (Avanzi, Simon, Goldman et Auchlin, 2010 pour les détails de l'étude). Bien entendu, les valeurs optimales des seuils que nous avions obtenues différaient pour chacun des genres composant la base de données C-PROM. Aussi, pour l'étude du corpus PROSO_FR, nous nous sommes servis des seuils suivants, obtenus après entraînement sur des échantillons de dix minutes de textes lus et de dix minutes d'indications d'itinéraires[5], soit, (i) pour le texte lu: hauteur relative = 1,43 dt ; Durée relative = x1,61; Glissando montant = 2,07 dt; (ii) pour la prescription d'itinéraire: hauteur relative = 2,6 dt ; Durée relative = x1,71 ; glissando montant = 2,47 dt. Les deux autres enregistrements du corpus PROSO_FR n'ont pas été exploités dans le cadre de cet article.

[4] Soit, du plus formel vers le moins formel: des lectures à haute voix, des discours politiques, des flashs d'information radiophoniques, des conférences universitaires, des interviews radiophoniques, des prescriptions d'itinéraires et des récits de vie.

[5] *Cf.* Avanzi (à par. b).

À noter enfin que pour qu'une syllabe soit détectée proéminente, il faut qu'au moins un des quatre paramètres soit activé. Dès lors, la syllabe apparaît en rouge dans la tire syllabique dupliquée juste en-dessous du tracé mélodique (*cf.* figure 2). On peut également visualiser les résultats de l'analyse automatique dans la tire « Pauto » (en dessous de la tire des syllabes). Pour chaque intervalle syllabique, un « p » est affiché si le seuil d'un seul des quatre paramètres est dépassé, un « P » est affiché si au moins deux paramètres sont activés. Dans l'énoncé analysé dans la figure 2, les syllabes *donne*, *cédé* et *pression* sont étiquetées « p » car le seuil de hauteur est le seul à avoir été activé. Quant aux syllabes *sentiment* et *ambiante*, elles sont notées « P » dans la mesure où leur identification a mobilisé au moins deux paramètres. L'intervalle reste vide si la syllabe n'a pas été détectée proéminente.

### *2.2 Catégorisation*

Une fois identifiées les proéminences, il reste à trouver un moyen pour estimer leur degré de saillance. La perception des proéminences n'est pas un phénomène binaire: dans une phrase donnée, une syllabe n'est pas proéminente ou non proéminente. Il y a différents degrés de proéminence (Terken et Hermes, 2000). Au moment où nous écrivons ces lignes, il n'existe pas de corpus annotés manuellement pour rendre compte de cette perception continue et permettant d'entraîner des systèmes d'apprentissage: aussi avons-nous dû choisir une autre façon de procéder afin de mettre au point un algorithme qui permette d'estimer le degré de saillance d'une syllabe donnée.

#### *2.2.1 Hypothèses*

Nous sommes partis de l'hypothèse suivante: *plus le nombre de paramètres acoustiques entrant en jeu dans l'identification des proéminences est important, et plus les seuils fixés sont dépassés, plus la proéminence est perçue*. Nous nous fondons sur l'hypothèse suivante: plus le locuteur mobilise les paramètres pour actualiser une saillance syllabique, plus il a des chances d'atteindre son objectif: faire comprendre à celui qui l'écoute l'importance perceptive de la syllabe en question (*cf.* le principe d'*effort code* de Gussenhoven, 2002). Par ailleurs, il existe de nombreux phénomènes de compensation entre les paramètres acoustiques qui participent à la mise en valeur des syllabes. Par exemple, un allongement syllabique peut être équivalent à un écart de hauteur sur le plan de la perception des proéminences (Rossi *et al.*, 1981; House, 1990). Il faut donc tenir compte de ces variations de stratégie de production dans les calculs.

#### *2.2.2 Algorithme*

En pratique, dans un premier temps, nous avons attribué une « note » à chacun des quatre critères utilisés pour statuer sur le caractère proéminent d'une syllabe (cf. §2.1.2. *supra*). Concernant les critères à seuil, cette note est calculée grâce à

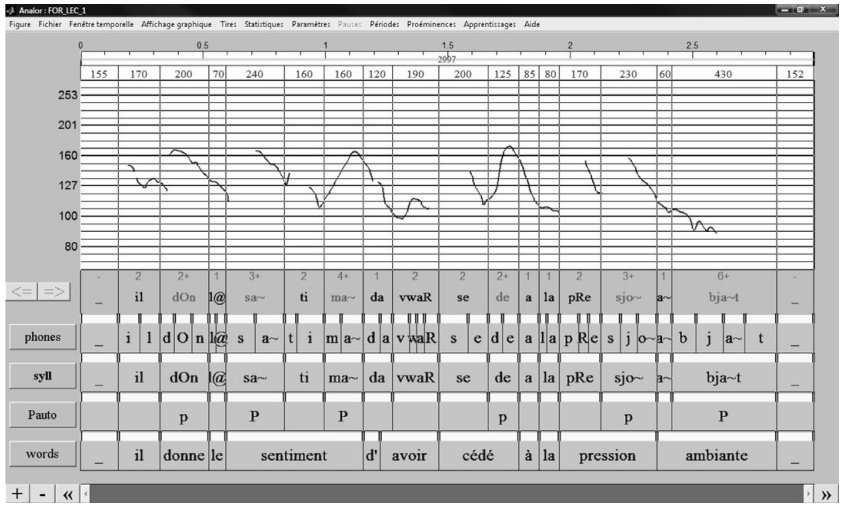


Figure 3. *Copie d'écran* ANALOR. *Illustration de la catégorisation des forces de proéminence. Analyse de l'énoncé: il donne le sentiment d'avoir cédé à la pression ambiante (*FOR-LEC*).*

« la fonction de proéminence » :

$$f(x) = 1/2 + 1/2 \cdot \tanh\left(2 \cdot \lambda \cdot \frac{x - s}{s}\right)$$

où $x$ est la valeur de la syllabe pour ce critère, $s$ le seuil pour ce critère, et $\lambda$ la pente de la fonction[6]. Pour le critère de pause adjacente, la note est de 10/10 en présence de pause adjacente, et de 0 sinon. La force de la syllabe est ensuite obtenue en faisant la moyenne pondérée des quatre notes obtenues, soit:

$$force = \frac{f_{durée}(x_{durée}).pds_{durée} + f_{hauteur}(x_{hauteur}).pds_{hauteur} + f_{montée}(x_{montée}).pds_{montée} + f_{pause}(x_{pause}).pds_{pause}}{pds_{durée} + pds_{hauteur} + pds_{montée} + pds_{pause}}$$

Ainsi, comme on peut le voir sur la figure 3, chacune des syllabes composant une phrase donnée se voit attribuer un score entre 0 et 10/10. On peut se figurer le détail de l'estimation du degré de saillance en cliquant sur les scores des syllabes, arrondi à l'unité près[7]. Une fenêtre apparaît[8], et affiche, de bas en haut: les valeurs et score de

[6] En faisant varier la pente, on rend cette fonction plus ou moins abrupte. Dans l'algorithme actuel, la pente de la fonction est de 1.5. Le poids des critères à seuil est de 1, alors que celui de pause est de 0.5.

[7] Les scores attribués aux syllabes qui n'ont pas été identifiées proéminentes lors de l'étape de détection ne sont pas à prendre en compte. Ils s'expliquent par le fait que certaines syllabes présentent des valeurs approchantes pour un ou plusieurs des seuils utilisés lors de la détection des proéminences (si une syllabe a une valeur de durée et de hauteur approchante, son score sera de 2).

[8] Pour des raisons de lisibilité, nous avons renoncé ici à faire apparaître cette fenêtre sur la figure 3.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | * |
| | | | | | | | | | | | | | | | * |
| | | | | | * | | | | | | | | | | * |
| | | | * | | * | | | | | | | | * | | * |
| | * | | * | | * | | | | * | | | | * | | * |
| * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * |
| il | dɔn | lə | sɑ̃ | ti | mɑ̃ | da | vwaʁ | se | de | a | la | pʁe | sjɔ̃ | ɑ̃ | bjɑ̃t |
| il | donne | le | sentiment | | | d'avoir | | cédé | | à | la | pression | | ambiante | |

Figure 4. *Grille métrique associée à l'énoncé* il donne le sentiment d'avoir cédé à la pression ambiante *(FOR-LEC), d'après la détection ANALOR.*

durée relative, de hauteur relative, de glissando montant et de pause. Pour la syllabe terminale du mot *sentiment*, ce sont les paramètres de hauteur et de glissando qui semblent jouer un rôle déterminant (5,48/10 et 8,48/10), la durée jouant un rôle moindre (0,99/10) et la syllabe n'étant pas suivie d'une pause silencieuse.

Pour la syllabe finale de l'énoncé *ambiante*, les valeurs de fo n'étant pas prises en compte (en l'état, le logiciel ne considère pas les valeurs d'écart négatives ou les configurations tonales descendantes comme des indices de proéminence, cf. §3.2.1. *infra* pour une discussion), c'est seulement en raison de la présence d'une pause silencieuse et d'un allongement de durée que l'on obtient une note de 6/10.

## 3. ILLUSTRATIONS SUR LE CORPUS

Dans cette dernière section, nous allons discuter de quelques exemples du corpus. Nous passerons en revue un cas où les prédictions faites par les règles phonologiques coïncident avec l'étiquetage proposé par le logiciel (§3.1), puis nous nous arrêterons sur trois autres exemples qui nous amènent à réfléchir sur la pertinence de certaines règles phonologiques existantes. Nous parlerons de la règle de dominance à droite (§3.2.1), de la règle du clash accentuel (§3.2.2) et de la règle des sept syllabes (§3.2.3).

### *3.1 L'étiquetage respecte les prédictions faites à partir des règles traditionnelles*

On peut, sur l'exemple dont nous nous sommes servis jusqu'à présent, montrer que la grille métrique que le logiciel a permis de mettre au jour est similaire en de nombreux points à celle que l'on a pu construire à l'aide des règles traditionnelles de l'intono-syntaxe. Comparer la grille de l'énoncé « *il donne le sentiment d'avoir cédé à la pression ambiante* » proposée dans la figure 1, avec celle de la figure 4 ci-dessous, construite à partir des résultats obtenue avec le logiciel (*cf.* figure 3 *supra*):

Les différences reposent sur la mise au jour d'un accent initial sur *sentiment*, qui n'était pas forcément prévu par les règles traditionnelles (mais dont l'occurrence

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | * |
| | | * | | | | | | | | | * |
| | | * | | | | | * | | | | * |
| | | * | | | * | | * | | | | * |
| * | * | * | * | * | * | * | * | * | * | * | * |
| a | vʁɛ | diʁ | sa | le | tɛ | de | ʒa | də | pɥi | lɔ̃ | tɑ̃ |
| à | vrai | dire | ça | l'était | | déjà | | depuis | | longtemps | |

Figure 5. *Grille métrique associée à l'énoncé* à vrai dire, ça l'était déjà depuis longtemps *(FOR-LEC)*.

n'est pas non plus prédite comme impossible à cet endroit-là); et sur le degré d'appréciation des forces des frontières terminales de groupe: le logiciel laisse à penser que la hiérarchie entre les groupes accentuels qui composent la phrase est en fait plus complexe que ce que les règles syntaxiques auraient pu avancer. D'après le découpage proposée par l'algorithme, il a y une sorte de rupture après le deuxième groupe accentuel (*il donne le sentiment // d'avoir cédé à la pression ambiante*), ce qui donne l'impression que la phrase est réalisée en deux syntagmes intonatifs distincts, et non en un seul. Cela dit, la grille proposée ci-dessus, plus proche de la réalisation effective, est parfaitement grammaticale.

### *3.2 L'étiquetage automatique ne respecte pas les règles de la phonologie prosodique*

Nous commenterons pour finir trois réalisations épinglées par le logiciel bloquées par les règles de la phonologie prosodique.

#### *3.2.1 Non-respect de la règle de dominance droite*

Dans les théories phonologiques du français, on trouve l'idée, formulée de diverses façons, que dans n'importe quel groupement de syllabes, c'est la syllabe la plus à droite qui porte la tête métrique du groupe, i.e. qui doit être la plus forte (Dell, 1984; Lacheret-Dujour et Beaugendre, 1999; Delais-Roussarie, 2005; cf. aussi le *Principe de Dominance Droite Systématique* chez Di Cristo, ce volume). Soit l'énoncé et sa grille métrique présentée dans la figure 5:

Selon les règles de la métrique standard, cet énoncé se réalise en deux groupes intonatifs. Le premier syntagme (*à vrai dire*) étant un adjoint à la phrase matrice (*ça l'était depuis longtemps*), on serait en droit de penser qu'il génère une frontière prosodique forte. On s'attendrait également à ce que la syllabe terminale de l'énoncé soit plus forte que toutes les précédentes, en raison du principe de dominance à droite. Or, les résultats de l'analyse automatique montrent autre chose:

Si le logiciel détecte bien une frontière forte sur le premier constituant, il ne détecte pas une frontière plus forte sur la syllabe terminale de groupe. Cette configuration métrique est « normale » dans les énoncés articulés en focus/

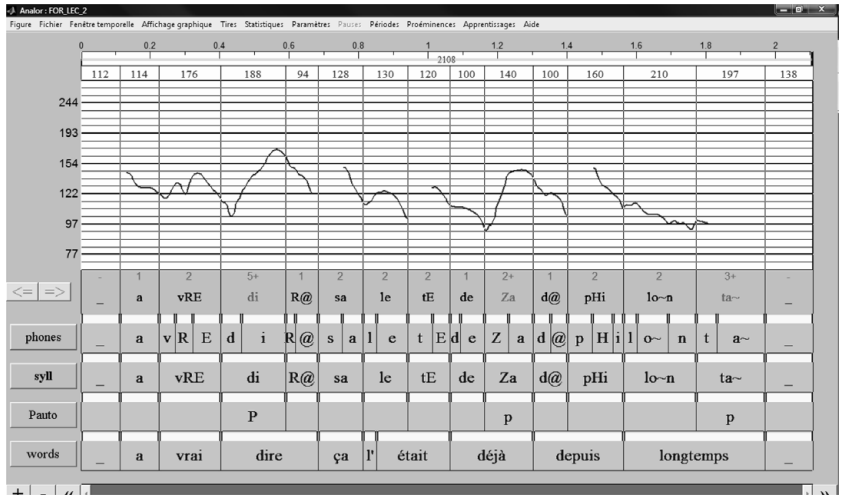


Figure 6. *Analyse par* ANALOR *de l'énoncé* à vrai dire, ça l'était déjà depuis longtemps *(*FOR-LEC*).*

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | * | | | | | | | | | |
| | | * | | | | | | | | | |
| | | * | | | | | * | | | | * |
| | | * | | | * | | * | | | | * |
| * | * | * | * | * | * | * | * | * | * | * | * |
| a | vʁɛ | diʁ | sa | le | tɛ | de | ʒa | də | pɥi | lɔ̃ | tɑ̃ |
| à | vrai | dire | ça | l'était | | déjà | | depuis | | longtemps | |

Figure 7. *Grille métrique associée à l'énoncé* à vrai dire, ça l'était déjà depuis longtemps *(*FOR_LEC*), d'après la détection* ANALOR*.*

post-focus, tels que les clivées: *c'est toi / qui le demandes* (FOR_DIA), où l'accent nucléaire, c'est-à-dire l'accent le plus fort de l'énoncé, frappe la syllabe *toi*, la fin de l'énoncé étant prononcée avec une intonation plate et peu modulée, typique des appendices (Delais-Roussarie, 2005; Dehé, 2009). Or, ce n'est pas un problème de portée de focus ici: le constituant *à vrai dire* a une valeur thématique et non rhématique. Le problème vient du fait que la syllabe finale de *longtemps* ne présente pas un allongement significatif, propre aux tons de frontière descendant à l'infra-bas. En outre, l'absence de modulation de fo est traitée par le logiciel comme une absence de proéminence.[9]

[9] Cette interprétation est conforme aux pratiques des auditeurs, qui ne voient pas dans les tons terminaux descendants de proéminences (Avanzi, à par. b). Nous ne développons

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | * |
| | | | | | * | | | | * |
| (*) | | * | | | * | | | | * |
| * | * | * | * | * | * | * | * | * | * |
| sə | fy | jɛʁ | dɑ̃ | lə | syd | ɑ̃ | na | vi | ɲɔ̃ |
| ce | fut | hier | dans | le | sud | en | Avignon | | |

Figure 8. *Grille métrique associée à l'énoncé* ce fut hier dans le sud en Avignon *(FOR-LEC)*.

*3.2.2 Non-collision accentuelle*

Une autre contrainte, bien connue des spécialistes, stipule que deux syllabes adjacentes, appartenant à deux mots lexicaux monosyllabiques distincts, ne peuvent être simultanément proéminentes. C'est la règle dite d'évitement de collision accentuelle (Garde, 1968). Soit l'exemple suivant, et sa grille métrique (figure 8):

On serait en droit d'attendre un accent sur *fut*, puisqu'il s'agit d'un mot plein, et donc générateur de groupe accentuel. Cependant, le fait qu'il soit suivi d'un adverbe monosyllabique en position finale de syntagme (*hier*, prononcé en une seule syllabe) entraîne sa désaccentuation. Ce mécanisme permet d'éviter que deux syllabes contiguës appartenant au même groupe soient toutes les deux proéminentes, pour éviter une collision d'accent. Cette résolution peut aussi prendre la forme d'un recul d'accent: auquel cas, la réalisation d'un accent secondaire sur le clitique *ce* pourrait être interprétée comme une réalisation « anticipée » de l'accent primaire attendu sur le verbe.

La détection automatique des proéminences permet de se rendre compte que dans la façon dont cet énoncé a été effectivement prononcé, la règle de non-collision accentuelle est enfreinte, puisque deux syllabes adjacentes (*fut* et *hier*) sont détectées proéminentes.

*3.2.3 Règle des sept syllabes*

Un dernier exemple nous permettra d'illustrer une autre règle traditionnelle dont les modèles phonologiques font état mais que les locuteurs peuvent transgresser. Soit l'énoncé de la figure 11:

La figure 11 donne une représentation de la structure métrique que n'importe quel modèle phonologique serait à même de prédire. Après analyse dans le logiciel (figures 12 et 13), on constate que si les constituants syntaxiques majeurs sont bien actualisés par des proéminences fortes (*là qui part euh de Nef Chavant, là le boulevard qui passe à côté d'Habitat*), on observe une tendance très nette à la

pas ce point ici, faute de place, à savoir le problème majeur dans l'analyse de la structure prosodique du français, qui concerne la distinction entre proéminence et frontière.

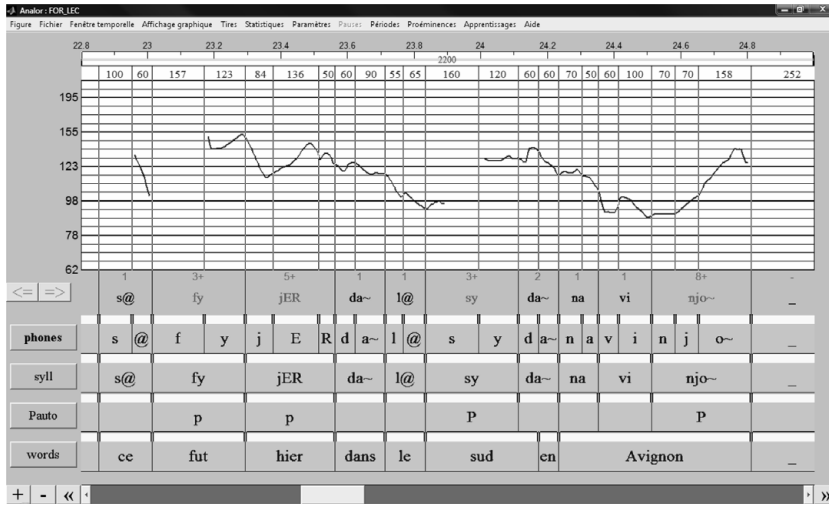


Figure 9. *Analyse par* ANALOR *à l'énoncé* ce fut hier dans le sud en Avignon *(*FOR_LEC*).*

| | | | | | | | | | (.) |
|---|---|---|---|---|---|---|---|---|---|
| | | * | | | | | | | * |
| | | * | | | | | | | * |
| | * | * | | | * | | | | * |
| | * | * | | | * | | | | * |
| * | * | * | * | * | * | * | * | * | * |
| sə | fy | jɛʁ | dɑ̃ | lə | syd | ɑ̃ | na | vi | ɲɔ̃ |
| ce | fut | hier | dans | le | sud | en | Avignon | | |

Figure 10. *Grille métrique associée à l'énoncé* ce fut hier dans le sud en Avignon *(*FOR_LEC*), d'après la détection* ANALOR.

désaccentuation à l'intérieur des deux syntagmes intonatifs. Cette désaccentuation n'a rien d'exceptionnel dans le premier syntagme. Les locuteurs ne sont pas obligés de réaliser des frontières de groupes accentuels aux frontières des groupes syntaxiques à partir du moment où l'espace entre les deux syllabes n'excède pas sept syllabes (Wioland 1985; Martin, ce volume). Partant, la façon dont est intoné le deuxième groupe intonatif est beaucoup moins « standard » de ce point de vue, puisque l'espace entre les deux proéminences n'est pas de sept syllabes mais de dix syllabes.

Ce genre de phénomène (ce qu'on appelle dans la littérature un *lapse*), fréquent dans la parole spontanée (Avanzi, à par. a), invite à se demander s'il ne faudrait pas plutôt reformuler la contrainte en tenant compte de la contrainte du temps, ce qui

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | * |
| | | | | | | * | | | | | | | | | | | * |
| * | | | | | | * | | | | | | | | * | | | * |
| * | | * | | (*) | | * | | | * | | (*) | | | * | | | * |
| * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * |
| la | ki | paʁ | də | nɛf | ʃa | vɑ̃ | lal | bul | vaʁ | ki | pa | sa | ko | te | da | bi | ta |
| là | qui | part | de | Nef | Chavant | | là l' | boulevard | | qui | passe à | | côté | | d'Habitat | | |

Figure 11. *Grille métrique associée à l'énoncé* là qui part euh de Nef Chavant là le boulevard qui passe à côté d'Habitat *(Infor-Iti).*

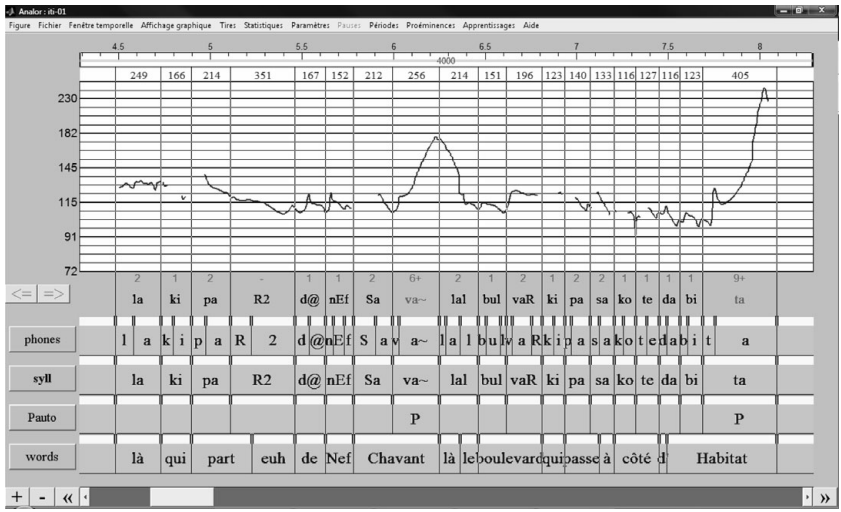


Figure 12. *Grille métrique associée à l'énoncé* là qui part euh de Nef Chavant là le boulevard qui passe à côté d'Habitat *(Extrait de Infor-Iti)*

| | | | | | | | | | | | | | | | | | (.) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | (.) | | | | | | | | | | | * |
| | | | | | | * | | | | | | | | | | | * |
| | | | | | | * | | | | | | | | | | | * |
| * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * |
| la | ki | paʁ | də | nɛf | ʃa | vɑ̃ | lal | bul | vaʁ | ki | pa | sa | ko | te | da | bi | ta |
| là | qui | part | de | Nef | Chavant | | là l' | boulevard | | qui | passe à | | côté | | d'Habitat | | |

Figure 13. *Grille métrique associée à l'énoncé* là qui part euh de Nef Chavant là le boulevard qui passe à côté d'Habitat *(Infor-Iti), d'après la détection Analor.*

reviendrait à prendre non pas en compte le nombre de syllabes brut, mais le taux d'articulation, calculé en fonction du débit moyen du locuteur.

## 4. CONCLUSION

Dans cet article, nous avons présenté un système qui permet de mettre au jour de façon semi-automatique la structure prosodique, plus précisément la structure rythmique, des énoncés rencontrés dans les corpus de français parlé. À partir d'une transcription alignée en phonèmes et syllabes, l'algorithme procède à une détection des saillances syllabiques, et en estime le degré de force en se basant sur les paramètres acoustiques classiques que sont la f0, la durée et la pause silencieuse. Nous avons ensuite illustré les résultats de cette détection en analysant et en commentant des exemples du corpus PROSO_FR (cf. Avanzi et Delais-Roussarie, ce volume). Nous avons montré que si le système pouvait donner des résultats très proches de ceux que les règles métrico-syntaxiques permettent de prédire, nous avons aussi mis en exergue que la détection automatique invitait d'une part à problématiser l'apparente synonymie entre frontière et proéminence (§3.2.1.), d'autre part à questionner la pertinence de certaines règles phonologiques traditionnellement envisagées comme inviolables (non-collision accentuelle, cf. §3.2.2.; règle des sept syllabes, cf. §3.2.3.; cf. Lacheret-Dujour & Beaugendre, 1999). L'intérêt d'un tel système, outre qu'il permet à des transcripteurs pas forcément experts de traiter de larges bases de données de façon rapide et cohérente, est qu'il offre un résultat relativement « objectif », puisqu'il se base uniquement sur les paramètres acoustiques, et ne fait pas entrer en ligne de compte le savoir phonologique que l'analyste a du système de sa langue. Au final, l'analyse de corpus de parole plus importants et plus variés avec un tel outil devrait permettre d'y voir plus clair dans la phonologie de la prosodie du français, et donc d'enrichir, en retour, les modèles formels existants.

## 5. REMERCIEMENTS

Les auteurs tiennent à remercier Corinne Astésano pour ses commentaires sur une version précédente de cet article, ainsi que les deux relecteurs anonymes de la revue JFLS. Cette recherche a bénéficié du soutien du Fonds National Suisse de la recherche scientifique (subsides n° PBNEP1–127788 et n°100012–126745, Université de Neuchâtel). Elle s'inscrit également dans le cadre de l'ANR Rhapsodie (ANR-07-CORP-030–01).

*Adresses pour correspondance:*
*Mathieu Avanzi*
*Chaire de linguistique française*
*Université de Neuchâtel*
*Ruelle Vaucher 22*
*2000 Neuchâtel*

*Suisse*
*e-mail: mathieu.avanzi@unine.ch*

*Anne Lacheret-Dujour*
*Université Paris Ouest*
*Bâtiment A - 408 A*
*200, avenue de la République*
*92001 Nanterre Cedex*
*France*
*e-mail: anne@lacheret.com*

*Nicolas Obin*
*IRCAM*
*1 place Stravinsky*
*75004 Paris*
*France*
*e-mail: nobin@ircam.fr*

*Bernard Victorri*
*Lattice-ENS*
*1 rue Maurice Arnoux*
*92120 Montrouge*
*France*
*e-mail: bernard.victorri@ens.fr*

BIBLIOGRAPHIE

Astésano, C. (2001). *Rythme et accentuation en français: invariance et variabilité stylistique*. Paris: l'Harmattan.

Avanzi, M. (à par. a). La dislocation à gauche en français parlé. Etude instrumentale. *Le français moderne*, 2011/2.

Avanzi, M. (à par. b). *L'interface prosodie/syntaxe en français parlé. Dislocations, incises et asyndètes.* Thèse de doctorat, Universités de Neuchâtel et de Paris Ouest Nanterre.

Avanzi, M., Lacheret-Dujour, A. et Victorri, B. (2008), ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure. *Proceedings of Speech Prosody'08, Campinas*, pp. 119–122.

Avanzi, M., Lacheret-Dujour, A. et Victorri, B. (2010). A corpus-based learning method for prominence detection in spontaneous speech, *Proceedings of Prosodic Prominence: Perceptual and Automatic Identification, Proc. Speech Prosody 2010 Workshop, Chicago, Illinois, May 10th*.

Avanzi, M., Simon, A. C., Goldman, J.-P. et Auchlin, A. (2010). C-PROM. An annotated corpus for French prominence studies, *Proceedings of Prosodic Prominence: Perceptual and Automatic Identification, Speech Prosody 2010 Workshop, Chicago, Illinois, May 10th*.

Avanzi, M. et Delais-Roussarie, E. (2011). Regards croisés sur la prosodie du français: des données à la modélisation. *Journal of French Language Studies*, 21/1.

Dehé, N. (2009). Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics*, 45/3: 569–615.

Delais-Roussarie, E. (2005). *Phonologie et Grammaire: Études et modélisation des interfaces prosodiques*. Mémoire d'Habilitation à Diriger des Recherches (HDR), Université de Toulouse-le Mirail.

Delais-Roussarie, E. (2008). Corpus et données en prosodie et en phonologie postlexicale: forme et statut. *Langages*, 171: 60–76.

Delais-Roussarie, E. et Post, B. (2008). Unités prosodiques et grammaire de l'intonation: vers une nouvelle approche. *Actes des Journées d'étude sur la Parole JEP-TALN 08*, Avignon, Juin 2008.

Delais-Roussarie et Yoo, H. (2011). Transcrire la prosodie: un préalable à l'échange et l'analyse de données. *Journal of French Language Studies*, 21/1.

Delattre, P. (1938). L'accent final en français: accent d'intensité, accent de hauteur, accent de durée. *French Review*, 12/2: 141–145.

Dell, F. (1984). L'accentuation dans les phrases en français. Dans: F. Dell, D. Hirst et J.-R. Vergnaud (dir.), *Forme sonore du langage: Structure des représentations en phonologie*. Paris: Hermann, pp. 65–122.

Di Cristo, A. (2011). Une approche intégrative des relations de l'accentuation au phrasé prosodique du français. *Journal of French Language Studies*, 21/1.

Garde, P. (1968). *L'accent*. Paris: Presses Universitaires de France.

Goldman, J. Ph., Avanzi, M., Lacheret-Dujour, A., Simon, A.-C. et Auchlin, A. (2007). A methodology for the automatic detection of perceived Prominent syllables in spoken French. *Proceedings of Interspeech'07*, pp. 91–120.

Guaïtella, I. (1997). Parole spontanée et lecture oralisée: activités cognitives différentes, organisations rythmiques différentes. *Travaux de l'Institut de Phonétique d'Aix*, 17: 9–30.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and Phonology. *Proceedings of the Speech Prosody 2002, Aix-en-Provence*, pp. 47–57.

Halle, M. et Vergnaud, J. R. (1987). *An Essay on Stress*. Cambridge, MA: MIT Press.

Hart, J. 't, Collier, R. et Cohen, A. (1990). *A Perceptual Study of Intonation: an Experimental-Phonetic Approach*. Cambridge: University Press.

Hirst, D. et Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15: 71–85.

House, D. (1990). *Tonal Perception in Speech*. Lund: University Press.

Lacheret-Dujour, A. et Beaugendre, F. (1999). *La Prosodie du français*. Paris: CNRS Editions.

Lacheret-Dujour, A. (2003). *La Prosodie des circonstants*. Louvain: Peeters.

Martin, Ph. (2011). La prosodie du français. Une approche pas très syntaxique. *Journal of French Language Studies*, 21/1.

Mertens, P. (2004). Le Prosogramme: une transcription semi-automatique de la prosodie. *Cahiers de l'Institut de Linguistique de Louvain*, 30.1–3: 7–25.

Mertens, P. et Alessandro d', Ch. (1995). Pitch contour stylization using a tonal perception model. *Proc. Int. Congr. Phonetic Sciences*, 13/4: 228–231.

Obin, N., Rodet, X. et Lacheret-Dujour, A. (2008). Un modèle de durées des syllabes fondé sur leurs propriétés intrinsèques et les variations locales de débit. *Actes des 27èmes Journées d'Etude sur la parole, juin 2008, Avignon*.

Portes, C. et Bertrand, R. (2011). Permanence et variation des unités prosodiques dans le discours et l'interaction. *Journal of French Language Studies*, 21/1.

Prince, A. (1983). Relating to the grid. *Linguistic Inquiry*, 14: 19–100.
Rietveld, T. et Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13: 299–308.
Rossi, M., Di Cristo, A., Hirst, D., Martin, P. et Nishinuma, T. (dir.). (1981). *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck.
Selkirk, E. (2005). Comments on intonational phrasing in English. Dans: S. Frota, M. Vigário et M.-J. Freitas (dir.), *Prosodies. With special reference to Iberian languages*. Berlin/New-York: Mouton de Gruyter, pp. 11–58.
Simon, A. C. (2004). *La structuration prosodique du discours en français. Une approche multidimensionnelle et expérientielle*. Berne: Peter Lang.
Terken, J. et Hermes, D. (2000). The perception of prosodic prominence. Dans: M. Horne (dir.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce*. Dordrecht: Kluwer, pp. 89–127.
Wioland, F. (1985). *Les Structures rythmiques du français*, Paris: Slatkine-Champion.